

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 9, 2000, 10:18:13 ; Search time 4054.2 Seconds
(without alignments)
3971.996 Million cell updates/sec

Title: US-09-117-447-1

Perfect score: 3687
Sequence: 1 atgagtagaanaaaagctgt.....ttacattctatataagtaa 3687

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1033670 seqs, 2183789903 residues

Total number of hits satisfying chosen parameters: 2067340

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

GenEmbl:*
1: gb_ba1:*
2: gb_ba2:*
3: gb_cm:*
4: gb_ov:*
5: gb_pat:*
6: gb_ph:*
7: gb_pl1:*
8: gb_pl2:*
9: gb_pl1:*
10: gb_pl2:*
11: gb_pl3:*
12: gb_ro:*
13: gb_sy:*
14: gb_un:*
15: em_fun:*
16: em_hum1:*
17: em_hum2:*
18: em_in:*
19: em_om:*
20: em_or:*
21: em_ov:*
22: em_pat:*
23: em_ph:*
24: em_pl:*
25: em_ro:*
26: em_sts:*
27: em_sy:*
28: em_un:*
29: em_vl:*
30: gb_hgt1:*
31: gb_hgt2:*
32: gb_in1:*
33: gb_in2:*
34: em_ba2:*
35: em_hum3:*
36: em_hum4:*
37: em_hum4:*
38: gb_pl4:*
39: gb_hgt3:*
40: gb_hgt4:*
41: gb_hgt5:*
42: gb_hgt6:*
43: gb_hgt7:*

44: em_hgt1:*
45: em_hgt2:*
46: em_hgt3:*
47: em_hum5:*
48: gb_pl3:*
49: gb_pl5:*
50: gb_hgt8:*
51: gb_hgt9:*
52: gb_hgt10:*
53: gb_hgt11:*
54: gb_hgt12:*
55: gb_hgt13:*
56: gb_hgt14:*
57: gb_in3:*
58: gb_hgt15:*
59: gb_hgt16:*
60: gb_hgt17:*
61: em_hgt4:*
62: em_hgt5:*
63: em_hgt6:*
64: em_hgt7:*
65: em_hum6:*
66: gb_hgt18:*
67: gb_hgt19:*
68: gb_hgt20:*
69: gb_hgt21:*
70: gb_hgt22:*
71: gb_hgt23:*
72: gb_vl1:*
73: gb_vl2:*
74: gb_ba3:*
75: em_hgt8:*
76: em_hgt9:*
77: em_hgt10:*
78: em_hgt11:*
79: em_hgt12:*
80: em_hgt13:*
81: em_hgt14:*
82: em_hgt15:*
83: em_hgt16:*
84: em_hgt17:*
85: em_hgt18:*
86: em_hgt19:*
87: em_hgt20:*
88: em_hgt21:*
89: em_hgt22:*
90: em_hgt23:*
91: gb_pl6:*
92: gb_pl7:*
93: gb_sts1:*
94: gb_sts2:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3687	100.0	3687	1 AX000218	AX000218 Sequence
2	3687	100.0	3687	2 A93832	A93832 Sequence 1
3	3687	100.0	3832	5 BSSBSA	X71092 B. Stearother
4	513	13.9	3680	1 AF055578	AF055578 Bacillus
5	487.6	13.2	3090	1 AF228338	AF228338 Bacillus
6	77.4	2.1	2069	22 E10125	E10125 DNA encodin
7	77.4	2.1	3399	22 E10126	E10126 DNA encodin
8	77	2.1	11154	8 AP002460	AP002460 Arabidops
9	73	2.0	5420	33 AF056936	AF056936 Plasmodi
10	71.4	1.3	17953	52 AC024253	AC024253 Homo sapi
11	64.2	1.7	110000	71 PFMALP1_1	Continuation (2 of
12	64	1.7	3720	73 S76368	S76368 ORF 5' of E

[illegible]

ALIGNMENTS

[illegible]

```

RESULT 1
AX0000218
LOCUS AX0000218 3687 bp DNA BCT 10-MAR-2000
DEFINITION Sequence 1 from Patent WO9906567.
ACCESSION AX000218
VERSION AX000218.1 GI:7240682
KEYWORDS
SOURCE
ORGANISM
.
Bacillus stearothermophilus.
Bacillus stearothermophilus
Bacteria: Firmicutes; Bacillus/Clostridium group;
Bacillus/staphylococcus group; Bacillus.
1 (bases 1 to 3687).
Lubitz, W. and Resch, S.
SECRETION OF CARRIER-BONDED PROTEINS INTO THE PERIPLASMA AND THE
EXTRACELLULAR SPACE
Patent: WO 9906567-A 11-FEB-1999;
LUBITZ WERNER (AT); RESCH STEPHANIE (DE)
FEATURES
source
1..3687
/organism="Bacillus stearothermophilus"
/strain="PV72"
/db_xref="taxon:1422"
/clone="SB5A"
1..3687
1..3687
/note="unnamed protein product"
/codon_start=1
/transl_table=1
/protein_id="CAB77069.1"
/db_xref="GI:7240683"
/translation="MDRKAVKLITASAIAASAFVAAPNPNSAATDVAIVVSOAKQA
FKKAYVYSHVTEGTEGEPNDVYAEVKNKKRPAVALYNRGAAGKAKAYADLDC
KEYETVYFKANPCKSGEARVATYIDATINATLDEMROLEAAYOAKDLERAEQYTHKTK
PELEIKRIVLIDRYIGKTRDLNSTFEAKQDLNDSLTIDITVAMKAREVQDAVKGK
NLDKAAVADQINDIYLPKVTDAEFTLEVAKKDLADAEALTRVESVSAINTQNQNR

```

[illegible]

QY 721 ttcaaaactgtaacagaaagtaagcgaanaaaagcattagatgcagatgaagctgcgctt 780
|||||
Db 721 TTCAAAGCTGAAGCTAACAGAAAGTAAAGGAAAGGATTTAGATGACAGATGAAGCTGGCGTT 780
QY 781 actccaaaagtctgaaagtgttaagtcgcatctaacactcaaaacaagctgtttaataaca 840
|||||
Db 781 ACTCCAAAAGTTGAAAGTGTAAGTGGATTAACACCTCAAAACAAAGCTTTGATTAATACA 840
QY 841 ggaagacacagtgaaacggaacactaaattacacatttcgcgcgtgcgaataaggtaca 900
|||||
Db 841 GGAGTACCACTGAACGGAACACTTAAATTCACATTCAGCTGCTGCAAAATGAAGTACA 900
QY 901 gtaaacgtaaacactgcgtacgtatctataaagtgcagcgaataatccattgccttaac 960
|||||
Db 901 GTAACGTAATAATACGTAGCTAGCTATTAAGTGAAGCGATTAATCCATTCGCTTAT 960
QY 961 acgagcagatgcttcttctatctacagacggaanaaaactaactcgtgtgaagcttcaactca 1020
|||||
Db 961 ACGGCAAGTGTCTTCTTATCTACAGACGGAAGAAACATCATCTGATGCTTCACTCCA 1020
QY 1021 ttgcgaataatacaggaagataaagtaagtaagtaagtaagtaagtaagtaagtaagtaag 1080
|||||
Db 1021 TTCGAAATTAATACGAGATATAAGTATAGTTAAAGTATTAACAAAAATGCAAA 1080
QY 1081 gaatttaagaagatgcattcactcttcaagcttcgaatgcgtgtgaagcttcaactcaagtg 1140
|||||
Db 1081 GAATTTAAAGAGATGCATTCATCTTCAGCTTCGAAATGATGCTGTATTTACTCAAGTG 1140
QY 1141 ttggaactaatgttaacaaacaaacactctgttaacttcgagcaggtactcttgcagact 1200
|||||
Db 1141 TTTGGAACATTAATGTAAACAAACACCTTCTGTAACCTTGACAGAGGTACTTTCGACACT 1200
QY 1201 gacgactacttaacaggtatcttgaataagttgttagacccctgaaacttaacaactcgtg 1260
|||||
Db 1201 GACGATCTTTTACAGTACTATTTGATAGTGTGTTAGCACTGGAACCTTAAACACACTCG 1260
QY 1261 aacgttactatctacagatgtctgaacactgcgaacacgcattccgaatttcgactactctt 1320
|||||
Db 1261 AACGTACTATTACAGATGTGAAACGTAACGCGATTCGAAATGCTGATTTGCTTACTTCT 1320
QY 1321 ggtctcaacttaactaactaactgaacgaacggtgtgaactcgtgtgaatacaataaactt 1380
|||||
Db 1321 GGTCTTACATTAATTAATGCTTAAAGAAAGCGTGTAGTAACTGCTAAACATTAATAACTT 1380
QY 1381 gctatcaataatgttaaaacttaactgttcaaatgcgaagacttaacgaatgaagtttc 1440
|||||
Db 1381 GCTATCAATTAATGTTAAACATTAATGCTTAAATGCAATGCAAGCTTACGAGTTAGTGTTC 1440
QY 1441 actgcaaaacgcatcagcacaacactgtgtgtacgcgtccctactaactttaagttgttacaact 1500
|||||
Db 1441 ACTGCAAAACGCAATCAGCAACACTGTGCTACGCGCTCCACTACTTAAGTGTGTAACAT 1500
QY 1501 ttactactggttctctcttaacaaagtcttggtggttaaaacttcggctggtgtgtgaatga 1560
|||||
Db 1501 TTAATCTACGTTGCTTCTTCAACAAATGTTGGGTAAATGTGGCTGGTGTGATGA 1560
QY 1561 gctggaactatatactcgtctgttcaatcacaacagcttgcgtactgaatgaagcgaa 1620
|||||
Db 1561 GCTGGAACCTTATTAATCTGCTTCAATTCACAAACACTTGTGTAAGTTAGCGAA 1620
QY 1621 tctacttttagcgtataactgttlatagttgaaagaacttcgtgtacagttgttcttct 1680
|||||
Db 1621 TCTACTTTAGCTGATTAATCTTGTATTAAGTGAAGAAAGATTCGTGATTTGCTGCTT 1680
QY 1681 gaactaataataatgcagacgcttaaatgttaactttagtgcgaagaacggaacttaaa 1740
|||||
Db 1681 GAACCTAAATTAATATGCAACACCTTAAATGTGAACCTTGTAGTGCACCAACGCGACTTAA 1740
QY 1741 gaaataatacatatcaataatcaaaatttaaaaagcgttgaagtcgaataaggtattga 1800
|||||
Db 1741 GAAATTAATCATTAATCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1800

QY 1801 tttagcactgtttaaaggaanaacatalaagttcaaaactcaagacttaactgtctctaca 1860
|||||
Db 1801 TTAGCAGCTGTTAAAGAGAAACATATGATGTTCAAAACCTCAAGACTTAATCTCTCTACA 1860
QY 1861 gttattaggttaagctcttaaaatggtgcagctgtgatttaaaagttaactgaagcttaaga 1920
|||||
Db 1861 GTTATTAGGTGTAAGCTTAAATTAATGCGAGCGCTGATTTAAAGTAACTGAACCTCAAGA 1920
QY 1921 ttactgtgaagtcttcagagaatttaaatatcaatttaactgaacacacgcttcgggtaac 1980
|||||
Db 1921 TTTACTGTGAATCTCTCAGAGAAATTTAAATACATTTAATGCTACAAACGCTTCGGGTAC 1980
QY 1981 acaatcacatacgcgtcaagttgtctgtagtaaaagcgggtgcgaacttaactgtctctaca 2040
|||||
Db 1981 ACAATCACATACGTCAGAGTTGCTGTAGTAAAGGGGGTCAAACTTATCTCTCTACA 2040
QY 2041 gcaagtgaatcaatccagctagttgtgaaggggttactcgtgcgaagatggaacatacaaa 2100
|||||
Db 2041 GCAAGTGAATCATCTTCAGCTAGGTGTTGAAGGGGTACTGCTCAAGATGGAACATACAA 2100
QY 2101 gtgaagttgtctgtcaacaaactagaacgttaacgaaggtaacaattagttgtctggt 2160
|||||
Db 2101 GTGAAGTTGCTGTACCAATTAATTAACGTAACCAAGGTAACAAATTAATGATGTTGCT 2160
QY 2161 aaaggctgcacacgctcctgtttaaagatgctgcaaaatgcgaatactttaacactaact 2220
|||||
Db 2161 AAAGGTGCAACAGCTCCTGTTAAAGATGCTGCAAAATGCAAAATTTAGTCACTAATCTAT 2220
QY 2221 atctataacttaacactggaaggttaagaagcgttaacagcacaagcgttcaaaagtatct 2280
|||||
Db 2221 ACTATTAATTTTAAACGTAAGGTGTAAGGTGTAAGGTGTAAGGTGTAAGGTGTAAGGTG 2280
QY 2281 aaagtgtactttaaagacgctgactgaacttaacacacttaagcaagcttgaagtt 2340
|||||
Db 2281 AAAGTGTACTTTTAAAGACCTGATGCACTGATTAACCTTAACGACCTGATGCAAGT 2340
QY 2341 caaaaattcactacaaatttagcgaagaatlaaaactcttagtgcgttcttaagttggt 2400
|||||
Db 2341 CAAAATTCATTAATCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2400
QY 2401 ggcgaagttaactgtctggaagaatttaacaaacaggaatggtgtgaatgctgtactgtga 2460
|||||
Db 2401 GGCAGAGTAACTGCTGAGAAATTAACAAACAGGATGGGTGATGCTGCTGCTGCAAA 2460
QY 2461 actgtatcagttgtctcccaagaacagatgcaatgttgaagttaacagcgtctgtgttaca 2520
|||||
Db 2461 ACTGATCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2520
QY 2521 ttaactggtcttgaacataacgacaagaatgcgaatgtgcgtctgtgttagtataagttct 2580
|||||
Db 2521 TTAAGTGGCTTGAACAAATTAACGCAAAAGATGCAAAATTCGCTGCTGCTGCTGCTGCT 2580
QY 2581 tctactgtatgaattgtcgtgtatgtatgtatgtatgtatgtatgtatgtatgtatgtatgt 2640
|||||
Db 2581 TCTACTGATGAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2640
QY 2641 cgttacaacagctgtgagacacactgttagcttctgtgaaagcgtgcgtctgaacaaagtgt 2700
|||||
Db 2641 CGTTCACACAGCTGAGACACACTGATGCTTGTGAAAGCTGCTGCTGCTGCTGCTGCTGCT 2700
QY 2701 caaaacgctctcgtctgcatctcccaacaaagcactgtgaattgtatatacaactaagaactata 2760
|||||
Db 2701 CAAAACGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2760
QY 2761 gttgaattcaatgaaactgatttagcgggaagttaaaacttgagaacacgtgtgttaagat 2820
|||||
Db 2761 GTTGAATTCATTAATGAACCTGATTTAGCGGAAGTTAAACCTGGAACAACTGCTGTTTAAAGAT 2820
QY 2821 gcaagcaggtgaatggtgtgagctgtgactgttaacagcatttagagcgttctcaaaataatt 2880
|||||
Db 2821 GCAAGCAGGTAAAGCGGTAGCTGATCTGTAACAGCATTTAGAGGTTCTTCAATTAATTAAT 2880
QY 2881 gttactactcactcaagaattaaagcgtgttactactgttactgttactgttactgttactgt 2940
|||||

|||||
Db 2881 GATTTCACCCATCCCAAGATTAAGCTGTGACGTTTACTCTTAACAATTGACGGT 2940
QY 2941 GTGAGAGATAAGTAGGTGACACACATCTTAATACATCTGTTCAAGACTGATCT 3000
Db 2942 GTGAGAGATAAGTAGGTGACACACATCTTAATACATCTGTTCAAGACTGATCT 3000
QY 3001 GCGAATCCACAGCTTATCTTCAATCAGCATGCTGACGGTGTGACCTTGTCT 3060
Db 3061 GCGAATCCACAGCTTATCTTCAATCAGCATGCTGACGGTGTGACCTTGTCT 3060
QY 3061 AAAACAATTAACAATGAATTAAGTACAGGATCGATCCAAACCAATCACTCTTAAGAAG 3120
Db 3061 AAAACAATTAACAATGAATTAAGTACAGGATCGATCCAAACCAATCACTCTTAAGAAG 3120
QY 3121 GCTGACGAGAACTTCTTACTATTAATTAATGATAATGATAATTAATAAACA 3180
Db 3121 GCTGACGAGAACTTCTTACTATTAATTAATGATAATGATAATTAATAAACA 3180
QY 3181 TACAAATATGATTCACCAAGGTGTAACACTTGACGAGTTTACTCAATATGAGTTAGCA 3240
Db 3181 TACAAATATGATTCACCAAGGTGTAACACTTGACGAGTTTACTCAATATGAGTTAGCA 3240
QY 3241 GTTCAAAAGATTTTCAAACTGTGATGATGATGATGATGATGATGATGATGATGAT 3300
Db 3241 GTTCAAAAGATTTTCAAACTGTGATGATGATGATGATGATGATGATGATGATGAT 3300
QY 3301 TCTGTTGCTACTGACGAGAACTGATGATGATGATGATGATGATGATGATGATGAT 3360
Db 3301 TCTGTTGCTACTGACGAGAACTGATGATGATGATGATGATGATGATGATGATGAT 3360
QY 3361 AGCTATACCTCAGGATCTGACGAGAACTGATGATGATGATGATGATGATGATGATGAT 3420
Db 3361 AGCTATACCTCAGGATCTGACGAGAACTGATGATGATGATGATGATGATGATGATGAT 3420
QY 3421 CCAAGTTCCTCAATCTCAGAGAGGATGATGATGATGATGATGATGATGATGATGAT 3480
Db 3421 CCAAGTTCCTCAATCTCAGAGAGGATGATGATGATGATGATGATGATGATGATGAT 3480
QY 3481 AATATACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3540
Db 3481 AATATACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3540
QY 3541 GATGCGAGTGTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3600
Db 3541 GATGCGAGTGTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3600
QY 3601 AACAGCAAGATCTTAAAGATTTGTTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 3660
Db 3601 AACAGCAAGATCTTAAAGATTTGTTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 3660
QY 3661 GATACATACATCTTATTAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 3687
Db 3661 GATACATACATCTTATTAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 3687

RESULT 2
LOCUS A93832 3687 bp DNA PAT 22-JAN-2000
DEFINITION Sequence 1 from Patent WO9728263.
ACCESSION A93832
VERSION A93832.1 GI:6741963
KEYWORDS
SOURCE
ORGANISM
Bacillus stearothermophilus.
Bacillus stearothermophilus
Bacteria; Firmicutes; Bacillus/Clostridium group;
Bacillus/Staphylococcus group; Bacillus.
REFERENCE
AUTHORS 1 (bases 1 to 3687)
TITLE Lubitz W. and Sleytr U.
JOURNAL RECOMBINANT EXPRESSION OF S-LAYER PROTEINS
Patent: WO 9728263-A 07-AUG-1997;
LUBITZ WERNER (AT); SLEYTR UWE (AT)
FEATURES
Location/Qualifiers

source
1. 3687
/organism="Bacillus stearothermophilus"
/strain="PV72"
/db_xref="taxon:1422"
/clone="SBSA"
1. 90
1. 3687
/note="unnamed protein product"
/codon_start=1
/transl_table=1
/protein_id="CAB69613.1"
/db_xref="GI:6741964"
/translation="MDRKKAVKLTATASIAASAVFAANPNASEATDVATVVSQAKQ
EKRAYVTSHTVEGEPNINDVAYAEKNEKKRRIDPVALVNGAGKKRAYLADQ
KEEYVYKRNPKSGEAPVATYDAVYATNLDPMROELBAVOAKOLEKROVYHKI
PEIKTRVYILDRIYKTRRLDLSSTRKQOEIRDSLIITYTAMRAREQDVKAKG
NLDRAKAVNDQINQILPRVIDAFTELTEVAKKLDADDEALITRVESSAINTONKA
VELAVPANGTILKQLSAAANEDEVNNTVAVIYVDGNIIPALNTADVSLSTDKTIT
VDASTPEENNTERYVYVYKIKDKNGKEKEDEAFEFKLNDAVNTQVEGTNTNTSVN
LAAGFEDDDTLVYFDEKLAPETVNSNNTITDVEGKRIPVIASTGSLITLKE
ALVYKOKKLIANNKLTGTYNARAYELVFNASAPVATAPPTIGTSTSLIT
NWKGLAGVNEACTYPRGLOFTTTRATKIDESTLADNFVYVEKESGYVASELKRYN
DAKNTLVPRADLENTIYQIKIRKGLSDKIEGTVNEKTYEFKTODLAPTVISY
TSKNGDAELKYTEAEQEFYKSENLNFNATVSGSTITYGQAVVYKAGANLSALTAS
DIIPASVEAVTGODGYKVKYANQLENNQYKLVFGKATGATVKAADKANNTLATNY
IYFTTEGODVTAFTVYKFGKDSLKDADAVTTLNDAGOKFTIOFSEELKSSSL
VGVKTVYKILNNGMVDAGTGTYSVAPKTPNANKRYAAVYTLGLDNMDKRLRV
YDKSSTGDIADVAGNVIKEDILIRYNSKRTVSVYRAADKQDNASAPFTSTALD
TTKSLVEFNEFDLAEPENIVVVDAGNANVACTVIALDSTNKFPYPSOEKLAGT
VYSVILDGVNRKVENTLSKYITSEKTVSANPTLSISLADGAVNDRKRTTIEFSDS
VPNPITILKAKDGTSTFVYLVNNEKRYKIVFHKGVLDEFTOYELAVSKFOIG
TDISKVTFIIGSVATDEVRKALVGVSGMNGSTODPAATRLRSVADFAVEPALOF
SEGIDLTNATVTYVNTDDKTEVVISKESVADADDAKTKETLVINVTPLVDSSTY
KRIYVSGVKAADAGVADITPTYIK"
mat_peptide
BASE COUNT 1316 a 660 c 715 g 996 t
ORIGIN
Query Match 100.0%; Score 3687; DB 5; Length 3687;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 3687; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 atggatagaagaaagccttgaaactagaacagcagcgtatcttcagcagcgtcatctt 50
Db 1 atggatagaagaaagccttgaaactagaacagcagcgtatcttcagcagcgtcatctt 60
QY 61 gtgcgtgcaatcccaaacgcttcgaagcgtctacagatgtagcaaacagtagtaagccaa 120
Db 61 gtgcgtgcaatcccaaacgcttcgaagcgtctacagatgtagcaaacagtagtaagccaa 120
QY 121 gcaaaaagccagcttcaaaaagcctactactactacagcctacacgttaacggaactggt 180
Db 121 gcaaaaagccagcttcaaaaagcctactactactacagcctacacgttaacggaactggt 180
QY 181 gaattcccaacattacagatgtatagctgtagatacaacaacgcaaaaagcgaatacgtt 240
Db 181 gaattcccaacattacagatgtatagctgtagatacaacaacgcaaaaagcgaatacgtt 240
QY 241 gattgagtagcatttagtgaataaagcaggttgccgcaaaaagcagcgtactactagctgt 300
Db 241 gattgagtagcatttagtgaataaagcaggttgccgcaaaaagcagcgtactactagctgt 300
QY 301 ttacaaaagaataagaacttagcttcaaaagcaaacctcaaatcggcggaagctgtt 360
Db 301 ttacaaaagaataagaacttagcttcaaaagcaaacctcaaatcggcggaagctgtt 360
QY 361 gtagcaacttagatgcatcttaacactagaacaaaatagagaagaatggcgcaagag 420
Db 361 gtagcaacttagatgcatcttaacactagaacaaaatagagaagaatggcgcaagag 420
QY 421 ctgaagcgtcgtcttcaagcaaaagatttagaaaaagcgaacaaatactatcacaaaatt 480
Db 421 ctgaagcgtcgtcttcaagcaaaagatttagaaaaagcgaacaaatactatcacaaaatt 480

Db	421	CTAGAGCGCTGCTTCAACCAAAAGATTGAAAAAGCAGAACAAATCTCTCAAAAATT	480
Qy	481	ccttaagaaatlaaaccctgcacagcattcttagatcgctalatgyltaaaacactcgt	540
Db	481	CCTTATGAAATTAACCTCGCACAGTCATTTTATGATCGCTATATGGTAAACACTCGT	540
Qy	541	gatttactcgcctctacatttaagaacaaagcacaagaacttcgcgcagacttaattat	600
Db	541	GATTATACCTCGCTACATTTAAAGCAAAAGCAAGAACTTCGCGCACACTTAATTAT	600
Qy	601	gatttaccggttgcacatgaaacgcgcgaagtacaaagcctgtgaaagaagcaattta	660
Db	601	GATTATTACCGTTCGATGAAAGCGCGGAGATGCAAGACCTGTGTGAAGAGCAATTTA	660
Qy	661	gacaaagccaaagctgcgtctgtatcaaatcaatcacttaccaaagaagtaacagatgc	720
Db	661	GACAAAGCTAAAGCTGCTGTGTATCAAAATCAATCACTTACCAAAAGTAACAGATGCT	720
Qy	721	ttcaaaactgaaactaacagaaagtagcgaanaaagcatlagaatgcagatgagcttc	780
Db	721	TTCAAAACTGAAGCTAAACAAGTAGCGAAAAAGCATTTATATGCAGATGAAAGCTGCGCTT	780
Qy	781	actcacaaggttgaaggtttaagtgsgattacaacccaacaaagcctgtttgatttaaca	840
Db	781	ACTCCAAAAGTTGTAAGGTGAAGTGGATTAAACACTCAAAAACAAGCTGTTGAATTACA	840
Qy	841	gcaatcacccagtgaaacggaacactaaattacaactttcagctgcgtgcgaatgaaatata	900
Db	841	GCAGTACCCAGTGAACGGAACACTAAAAATTACAACCTTCAGCTGCTGCAAATGGAAGTACA	900
Qy	901	glaaacgttaatactgtatgcgtatctataaagtgagcgtatacatccatttgcccttaat	960
Db	901	GTLAAACGTAAATACCTGTAGTATCTATTAAGTGGACGGTAACTCATTTGCCCTTAAT	960
Qy	961	agggagaagatgttcctttactacagagcggaanaacataccctcgttgatgtttcaactcca	1020
Db	961	AGGGAGATGTTCTTTATCTACAGACGGAAAACTATCACTGTGGATGGCTTCAACTCCA	1020
Qy	1021	ttcgaanaataatacgcgaagataaagaatagtaagttaaaggtttaagacaaagaatgcaaa	1080
Db	1021	TTTCGAAAAATATACGAGATATTAAGTAGTAGTTTAAAGGTATTAAAGCAAAAAATGGCAAA	1080
Qy	1081	gaatttaaagaagatgcatcttcactccaagcttcgaatagatcgttaagttaactcaagt	1140
Db	1081	GAATTTAAAGAAATGATTCATCTTCAAGCTTCGAAGTATGATGCTGTAGTTACTCAAGTG	1140
Qy	1141	tttggaactcaaatgtaacaaacaaacacttctgttaacttagcagcaggttaacttcgcacact	1200
Db	1141	TTTGGAACTAAATGTAAACAACAACACTTGTGTAACCTTAGCAGCAGGTACTTTTCGACACT	1200
Qy	1201	gaacgaacttaacagtagtatattgataagttgttagcacctgaacactgaacacagctcg	1260
Db	1201	GACGATACTTTAACAGTAAATTTTGTAAAGTTGTTAGCACCTGAAACTGTAAACAGCTCG	1260
Qy	1261	aacgttactattacagatgttgaaacttgzaaazacgcattccagtaattgcatctactctc	1320
Db	1261	AACGTTACTATTACAGATGTTGAAACTGGAACACGCAATTCAGATAATGTGACTACTCTCT	1320
Qy	1321	ggtttctacacataactcattcgtttaaagaagcgttaagtaactggtataacacataaactt	1380
Db	1321	GGTTCTACAAATTACTATTTCGTTAAAGAAAGCGTTAGTAACCTGGTAAACATTAATTAACCTT	1380
Qy	1381	gctacacaataatgtaaaacacataaactggtttacaatgcagaagccttaacgaattagtgctc	1440
Db	1381	GCTACAAATATGTTTAAACATTAATTAACCTGTTACAAATGCAGAACTTACAGATTAGTGTTTC	1440
Qy	1441	actgcaaacgcacgcacacaaactgttgtaacgcctcacttaactttaagttgtataaact	1500
Db	1441	ACTGCAAAACGCATACGACCACTGTTGTACACGCTCTACTACTTTAGTGTGTACAACT	1500
Qy	1501	ttatctactcgtgtcctcttacaacaaatgttttgggttaaatgttcgtgtgtgtgtgaatgaa	1560
Db	1501	TTATCTACTCGTTCCTTTAACAAATGTTTGGGTAAATTTGGCTGTGTGTGTGAATGAA	1560
Qy	1561	gctggaacttattccttggtlcttcaatccaacaaacglttgctactaagttagacgaa	1620
Db	1561	GCTGGAACCTATTATCCTGCTCTCAATTACAAACAACGTTTGCTACTAAGTTAGAGAA	1620
Qy	1621	tctactttagctgataactttgttatctgttgaaaagaactcgtgtataagttgtgtcctc	1680
Db	1621	TCTACTTTACTGATAACTTGTGATTGATTGTAAGAAAAAATCTGGTAAAGTTGTGTCTCT	1680
Qy	1681	gaactaaacataatgacagcgcgaacaaatgtaactttagtgcgaagcgcgaactcaaa	1740
Db	1681	GAACTTAAATATATGACAGCGCTTAAATGTGAATCTTATAGTCCAAAAAGCGACCTTAAA	1740
Qy	1741	gaaataacacatctatcaatcaacaaatlaaaagccttggaagtcogaataaagttatgaa	1800
Db	1741	GAAATATCAATCTATCAATCAAAATTAATAAAGAGCTGGAAGTCCGATTAAGATTAGAA	1800
Qy	1801	ttaggcactgtttaacgagaacacataatgattccaacactcaagaacttaactgtcctaca	1860
Db	1801	TTAGGCACCTGTTAAACGAAAACTAATGAAGTTCAAACTCAAGCTTAACCTGCTCTACA	1860
Qy	1861	gttatatgcgttaacggttcaaaaagcgcgcagcttgatctaaagtaacttgaaagcccaagaa	1920
Db	1861	GTTATTAGCTTAACGTTTAAATATGGGACGCTGGATTTAAAGTAAGTAAGTCAAGCTCAAGAA	1920
Qy	1921	tttactgtgaagttctcagaagaatttaatacatltaacttcaaacacglttcggtgtagc	1980
Db	1921	TTTACTGTGAAGTTCTCAGAGAATTTAAATATCAATTTATGCTACAAACGTTTGGGTTAGC	1980
Qy	1981	acaatcacatacgcgtcaagttgcgtatgaataaagcgggtgcacaaactatactgctctaca	2040
Db	1981	ACATACATACAGGTCAAGTGTGCTGTGTAAAGCGGGTGCAAACTTATCGCTCTTACA	2040
Qy	2041	gcaagtgacatacttccagtaagtgttgaagcgttatcgttgtaagttgaaactataaazaa	2100
Db	2041	GCAAGTGACATCATCTCCAGCTAGTGTGTAAGCGGTTACTGTGTCAGATGGAACATACAAA	2100
Qy	2101	gtgaagtttgctgtaacacattagaagcgaacaaaggtacaaatagtagttgttcggt	2160
Db	2101	GTGAAGTTGCTGCTGTAACCAATTATGAACGTATACCAAGGTATAGTAGTGTTCGGT	2160
Qy	2161	aaagtgcaaacagctccctgtttaagaatgctgcgaatgcgaataacttagcaactaactat	2220
Db	2161	AAAGTGCAACACACTCCTGTTAAAGATGCTGCAAAATCAAAATCTTTAGCAACTACTAT	2220
Qy	2221	atctatacatttaacacttgaaggttcaagaagtaacgcacccaacggtttaacaaagtatctc	2280
Db	2221	ATCTATACATTTACATCACTGAAGGTCAAGAGCTAACACACCAACGGTTTACAAAAGTATTC	2280
Qy	2281	aaagtgatctctttaaagaacgcgtgatgcagttactacactttagcaacggttgatgcaggt	2340
Db	2281	AAAGTGATCTCTTTAAAGACGCTGATGCGATTACTACACTTAGCAACGTTGATGCGAGGT	2340
Qy	2341	caaaaattcacataccaactttagcgaagaattaaanaactctagttgttctttagtgggt	2400
Db	2341	CAAAAAATTCATCTATCCAAATTTAGCGAAGATTTAAAACTCTAGTGGTCTTTAGTGGGT	2400
Qy	2401		

Query Match 100.0%; Score 3687; DB 2; Length 3832;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 3687; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy 1 atggaagaagaaacgtgtgaaactagcaacagcaagtgctattgacgaagtgcatctt 60
Db 37 ATGGAAGAAAAAGCTGTGAACTAGCAACAGCAAGTGTATTGCAAGCAAGTCATTT 96
Qy 61 gtgcgtgcaaatccaaagccttcgaagcgtgtacagatgtgtgcaacagtagtaagcaa 120
Db 97 GTCGCTGCAAAATCCAAAGCCTTCTGAAGCGGCTACAGATGTAGCAACAGTACTAAGCCAA 156
Qy 121 gcaaaagcagcttcaaaaaagcactatattacttaccacacatagttacggaacttgt 180
Db 157 GCAAAAGCAGCTTAAAAAAGCACTACTACTTACAGCCATACGTAAACGAAACTGGT 216
Qy 181 gaattcccaaacattaaagatgtatattgtgaatacaacaaagcgaaaaaaacgatacgt 240
Db 217 GAATCCCAACATTAACGATGTATATGCTGAAATACAAACAAAGCGAAAAAGATACCGT 276
Qy 241 gatcggtgagcatatagtaataaagcaggtggtcgcaaaaaaaagcgttacttagtgtat 300
Db 277 GATGGGGTAGATTAAGTAAGCAAGAGTGGCGGAAAAAAGAGCGCTTACTTACCTGAT 336
Qy 301 ttacaaaagatatgaacttaagcttttcaaaagcaaacccctaaatctggcgaaagctgt 360
Db 337 TTACAAAAGATATGAAACTTACGTTTTCAAAGCAAAACCTTAAATCTGGCAACCTGCT 396
Qy 361 gtaagcaacttaccatcgatgtcttaccactatgcaacaaatagacgaataatgctgcgaag 420
Db 397 GTAGCAACTTACATGATGATGCTTACAAATGATGCAACAAATTAAGCAATATGGCCAAAG 456
Qy 421 ctgaagcgtgtgttcaagcaaaagattagaaaaaagcagaacaaactactatcaaaatt 480
Db 457 CTAGAGGCTGCTGTCAAGCAAAAGATTAGAAAAAGCAAAACAACTATATCACAAATT 516
Qy 481 ccttatgaataataaacctcgcaagcttctttagatcgctatagtggtcaaaacacttgt 540
Db 517 CCTATGAAATTTAAACCTGCGCAAGTCACTTTAGATCCGCTATATGTTAAACAACTGCT 576
Qy 541 gatttactgctctacatttaaaagcaaaagcacaagaactctgcgaagctttaaattat 600
Db 577 GATTACTTCCCTCTACATTAAAGCAAAAGCAAGCAAGCAACTTCGGACAGCTTAATTAT 636
Qy 601 gatattacggtgtgcaatgaagcgtgcgaagctacagaagcgtgtgtgaagcaggaactta 660
Db 637 GATATTACCGTTGCATGAAGCGCGCAAGTACAAAGCGCTGTAAAAAGCAGGCAATTTA 696
Qy 661 gacaaagctaaagctgtgtgtgatcaaatcaatcaatacttaccaaaagttaacagatgt 720
Db 697 GACAAAGCTAAAGCTGCTGTTGATCAAAATCAATCAATCTTCCAAAGTACACATGCT 756
Qy 721 ttcaaaactgaactaacaagaagtagcgaaaaaagcatatagatgcagatgaagctgcgtt 780
Db 757 TTCAAAACTGAACTACAGAGTAGTAGCAAAAAAGCATTAGATGAGATGAAGCTCGGCTT 816
Qy 781 actccaaaagtgtgaagtgtaagtgcatctaaactcaaaacaagaagctgttgaattaca 840
Db 817 ACTCCAAAAGTTGAAAGTGAAGTGCATTAACCTCAAAACNAAGCTGTTGAATTAACA 876
Qy 841 gcagtaacagtgaaagcgaacactaaatttacaacttcaagctgtgcgcaatgaagaatca 900
Db 877 GCAGTACCAAGTGAACGGAACACTTAAATTAACACTTTTCAGCTGTGCAAAAGTAAATACA 936
Qy 901 gtaaacgtlaaactgtatctataaagtgtgaacggttaacattccatttgccttatt 960
Db 937 GTAAGCGTAAATACTGTAGTATCTATAAGTGAAGCGGTAACATTCATTTGCCCTTAAT 996
Qy 961 acggcagaatgttcttctatctacagaacggaacaaactatcaactgtgtgagcttcaactca 1020
Db 997 ACGGCAAGATGTTCTTTATCTACAGACGCAAAACATATCATCTGTGTGATGCTTCAACTCA 1056
```

```
Qy 1021 ttcgaaataatagcaaglataaagtagttaaaggtattaaagacaaacaaatgycaaa 1080
Db 1057 TTGCAAAATATATACGAGATATTAAGTAGTACTTAAGGTATTAAGCAAAATGGCAAA 1116
Qy 1081 gaattaaagaagatgcatctcaacttccaagcttcgaatgtagtgcgttactcaagtg 1140
Db 1117 GAATTAAAGAAAGATGCTATTCATCTTCAAGCTTGAATATGATGCTGTAGTACTCAAGTG 1176
Qy 1141 ttggaactaatgttaacaaacaaacactctgttaacttagcagcaggaagctttgagact 1200
Db 1177 TTTGGAACATAATGTAAACAACACCTTCTGTAAACTTAAAGCAAGTATCTTGGACACT 1236
Qy 1201 gacgaactttaacagtagtatttgaatgttgttaagcactggaacagttaaacagctcg 1260
Db 1237 GACGATCTTTAACAAGTGTATTTGATAGTGTGTAGACCTGGAACCTGTAACACGCTCG 1296
Qy 1261 aacgttactatcacagatgttgaacatggaacacgcatctccaagttaatgcatctactct 1320
Db 1297 AACGTACTATTACAGATGTGTGAACCTGGAACGCAATTCAGATTAATGCAATCTACTTCT 1356
Qy 1321 ggttctacaattactattacgtttaaagaagcgttagtaactgtgttaacaatatcaact 1380
Db 1357 GGTCTTACAAATTAATTAATTAAGTAAAGAGCGTTAGTAACTGGTAAACAAATATAAATCTT 1416
Qy 1381 gctatcaataatgttaaaacattaacgtgttacaatgagaagcttaacagtagtgttc 1440
Db 1417 GCTATCAATATGTATAAACCTTAACGTGTACATGCAAGAGCTTACGAATTTAGTGTTC 1476
Qy 1441 actgcaaacgcatacagcaccacactgtgtctacgcgtccctactactttaggtgtgacaact 1500
Db 1477 ACTGCAAAACGATCAGCAGCAACATGTTGCTACCCCTCTACTACTTATAGTGTGATCAACT 1536
Qy 1501 ttacttactgttctcttacaacaaatgtgtgggttaaatggtcgtgtgtgtgtatgaa 1560
Db 1537 TTATCTTACTGCTGTCTTACCAACAAATGTTGGGTAAATGGCTGGTGGTGAATGA 1596
Qy 1561 gctggaacttatacctcgtgcttcaaatcacacaagaagtttgtaactaaagttagacgaa 1620
Db 1597 GCTGGAATTTATTCCTCGGCTTCAATTCACAAACAGTTTGTCTACTTAATTTAGACGA 1656
Qy 1621 tctactttagctgtaaacatttgatagttgaaaaaagaatctgtgtacagttgtgtctct 1680
Db 1657 TCTACTTACGCTGAATTACTTGTATAGTTGAAAAAGCAATCTGTATCAGTTGTTGCTTCT 1716
Qy 1681 gaacctaataatlaatgcgaagcgttaaaatgttaactttagtgccaaaaagcgaacttaaa 1740
Db 1717 GAACATAAATATAATGCAAGAGCTTAATAATGGTAACTTATAGCCAAAGCGGACCTTAAA 1776
Qy 1741 gaaaatacaacttatacaaatcaaatlaaaaaagcgttgaagtcgcgataaaggtatgaa 1800
Db 1777 GAAATATCAATCTATCAAAATTAATAAAGGCTTGAAGTCCGATTAAGGTATTGAA 1836
Qy 1801 tttagcactgtttaaagaaacacalataggttcaaaaactcaagaacttaactgtctctaca 1860
Db 1837 TTAGGCACTGTTAAGCAAAACATATGATGTTCAAAACTCAAGACTTAACGTCTCTACA 1896
Qy 1861 gttattagcgaagcttcaaaaatggtgcgaagcgtgtgattaaagtaactgcgaagctcaagaa 1920
Db 1897 GTTATTTACCGTGAAGCTTAAATAATGGCGACCTGTGATTAATAAGTAAGCTCAAGAA 1956
Qy 1921 tttaactgtgaagttctcagaagaatttaaatatcatcttaagtgtacaaacccgttcgggtgac 1980
Db 1957 TTTACTGTGAAGCTTCTCAGAAATTTTAATTAATCAATTAATGCTAACAACCGTTTCGGGTAGC 2016
Qy 1981 acaatcaatatacgggtcaagttgtctgttagtaaaagcgggtgcaaaacttatctgtcttaca 2040
Db 2017 ACAATCACATACGGTCAAGTGTGTGTAGTAATAACCGGCTGCAAACTATCTGTCTTACA 2076
Qy 2041 gcaagtgacatcatctcagcgttagtgttgaagcgggttactgtgtcaagatgtaaatacaaa 2100
Db 2077 GCAAGTGACATCATCTCCAGCTAGTGTGAACCGGCTTACTGTGTCAAGATGTAACATACAAA 2136
Qy 2101 gtgaagatgtgtcttaaccaattagaaggttaaccaaggttaaccaattagtagtgttgt 2160
```

Db	2137	GTGAAGTGTGCTGCACCAATTAGAACCAAGGGTACAAATTAAGTGTGGGT	2196
Qy	2161	aaaggtgcacacgctcctgttaagaatgctgcaaatgcaataacttaagcaactaactat	2220
Db	2197	AAAGGTGCACAGCTCCCTGTTAAAGATGCTGCAAAATGCAAAATCTTTACCACTTAACCTT	2256
Qy	2221	atctatacttacaactgaaggtcagaacgctaacagcacaacggttacaagaatctc	2280
Db	2257	ATCTATACATTTACAACTGAAGTGAAGCTTACAGCTACCAAGCTTCAAAAAGTATTC	2316
Qy	2281	aaaggtgactcttaaaaagacgctgagtgcaactacacttaacgaactgtagtcaggt	2340
Db	2317	AAAGGTGATTTCTTTAAAGACGCTGATGAGTACTTACACTTACGAACTGATGACAGGT	2376
Qy	2341	caaaaattcacatccaatttagcgagaagaataaaaactctcagtggtctcttaagtggt	2400
Db	2377	CAAAAATTCACATCCAAATTTAGCGAAGATTAATAAACTTTAGTGTCTTTAGTGGGT	2436
Qy	2401	ggcaaaagtacgtcgagaaatttaacaacaacgagtggtgtagtgctgtaactggaaca	2460
Db	2437	GGCAAAAGTACCTGTGAGAAATTAACAACAACGATGGGTGATGCTGTACTGGAACA	2496
Qy	2461	actgtaacgcttgctccctaagaacagatgcaaaatggtaaagtaacagcgtgctggttaca	2520
Db	2497	ACTGTATCAGTTGCTCCCTAAGACAGATGCNAATGTAAGTAAACAGCTGCTGTGTACA	2556
Qy	2521	ttaactgtgcttgacaataacgacaagaatggaatgtcgctgtagtaataagttct	2580
Db	2557	TTTAACTGGCTTGGACATACGACAAAGATGCAAAATTCGCTGTGATTAATTAAGTCT	2616
Qy	2581	tctactgtagaattgctgtagtgtagctgtagtaagttaataagaaagatacttaatt	2640
Db	2617	TCTACTGATGAGATGCTGATGATGATGATGATGATGATGATGATGATGATGATGAT	2676
Qy	2641	cgtaacaacgctgtagagacacactgtagctctgtagaagctgctgtagacaagaatggt	2700
Db	2677	CCTTCAACAGCTGGAGACACACTGTAGCTTCTGTGNAAGCTGCTGTACAAAGATGCT	2736
Qy	2701	caaaagcgtctgctgcaatcccaacaagaactgcaatgatacaactaagaagcttatta	2760
Db	2737	CAAAAAGCTTCTGTGCTCATCCCAACACACTGCAATGATGATGATGATGATGATGAT	2796
Qy	2761	gtgaattcaatgaactgatttagcgsaagtaaaactcgaagaacactgctgtttaaagt	2820
Db	2797	GTTGATTCATGAACTGATTTAGGGAAAGTTAAACCTGAGAACATGCTTTTAAAGAT	2856
Qy	2821	gcagcaggttaatgctgtagctgtagctgtagaagcattagaacggttctacaataaatt	2880
Db	2857	GCAGCAGGTAAATGCGGTACTGTACTGTACAGCATTTAGACGGTCTCAAAATAAATT	2916
Qy	2881	gtatcacctccatccaagaattaaagctgtgtacagtttactctgttaacaattgagct	2940
Db	2917	GATTTCACCTCCATCTCAAAATTAAGGTGGTACAGTTTACTCTGTAAATGACGGT	2976
Qy	2941	gtgagagataaagtagtgaacacaactcttaataactactcgttcaagaactgatat	3000
Db	2977	GTGAGAGATAAAGTAGGTAAACAAATCTTAATATATTACTTCCGTTCAAGACTGATCT	3036
Qy	3001	gcgaatccaagcttatcttcaatcagcaatgtgtagcagtgtagtaagctgtagcgttct	3060
Db	3037	GGGAATCCACAGCTTATCTTCAATCAGCATGCTGACGGTGTAACTTAACTTAACTTCT	3096
Qy	3061	aaaaacataaagttaaggttaggttagtgcacacccaacaaactcctttaaag	3120
Db	3097	AAAAACATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	3156
Qy	3121	gtcgaagcaactcattacttaactaactcattagtaataagtaataataaataaaca	3180
Db	3157	GCTGACGGAACCTCTTACTATTAACACTTTAGTAAATGTAAATTAAGAAAATTAACA	3216
Qy	3181	tacaaatgtatctccacaagaagtgtaacacttgacaggttactcaatatgattaga	3240

Db	3217	TACAAATTTGATTCACAAAGGTGTAACTGACGAGTTTACTCAATATAGATTACA	3276
Qy	3241	gttcaaaaagatttccaactggtacgatactgtagcaaaagttaacatcacaaggt	3300
Db	3277	GTTCAAAAGATTTTCAAACTGATGATGATGATGATGATGATGATGATGATGATGAT	3336
Qy	3301	tctgttctactgacgaagtaaaactcgtctcagtagcgttggttcaatggaatgaaca	3360
Db	3337	TCTGTGCTACTGACGAAGTAAACCTCTCTAGTAGGCGTGGTTCATGGAATGGAACA	3396
Qy	3361	agctactacgagatgctgcagcaacacgactcgtctgtagctcgttgctgag	3420
Db	3397	AGCTATCTCAGGATGCTGCAGCAACACAGCTTCTGCTGTAGTCACTGCTTGTCCGAG	3456
Qy	3421	ccagttgccttcaattctcagaagtagtgatgtaagaatgcaactgtagaataaca	3480
Db	3457	CGAGTGCCTTCAATTTCTCGAAGATTCGATTTAAAGAAATGCAACTGTGACATTAACA	3516
Qy	3481	aatactactgataaaactgltgaagtatttcaaaaagagagtgtagacgcagacat	3540
Db	3517	AATATTACTGATGATTAATAAAGCTTGAAGTTATTTCAAAAAGAGGTAGACGACACAT	3576
Qy	3541	gatgcaggtgctactaagagacattagtaataacacagttactcctttagtactg	3600
Db	3577	GATGCAGCTGCTACTAAGAGACATTAATTAACACAGTACTCTTGTGTACTTGTAT	3636
Qy	3601	aaccagcaagcttataaagattgttgtagtgagtgtagaagtagcagcaggaatgca	3660
Db	3637	AACAGCAAGCTTATTAATGTTGTGTAGTGAAGTAAAGATGACAGAGATTAATGTGCA	3696
Qy	3661	gatactatactcattataatgaagtaa	3687
Db	3697	GATCTATTACTTCTATATTAAGTAA	3723

RESULT	4			
AF055578	3680 bp	DNA	BCP	08-MAR-2000
LOCUS	AF055578	Bacillus stearothermophilus surface layer protein precursor (sbac)		
DEFINITION	AF055578	gene, complete cds.		
ACCESSION	AF055578			
VERSION	AF055578.2	GI:4581955		
KEYWORDS				
SOURCE				
ORGANISM		Bacillus stearothermophilus.		
REFERENCE		Bacillus stearothermophilus		
AUTHORS		Bacteria: Firmicutes; Bacillus/Clostridium group;		
TITLE		Bacillus/Staphylococcus group; Bacillus.		
JOURNML		1 (bases 1 to 3680)		
REFERENCE		Jarosch,M., Egelseer,E.M., Mattanovich,D., Sleytr,U.B. and Sara,M.		
AUTHORS		S-layer gene sbac of Bacillus stearothermophilus ATCC 12980:		
TITLE		molecular characterization and heterologous expression in		
JOURNML		Escherichia coli		
REFERENCE		Microbiology 146 (Pt 2), 273-281 (2000)		
AUTHORS		20170659		
TITLE		2 (bases 273 to 3572)		
JOURNML		Jarosch,M., Egelseer,E.M., Mattanovich,D., Sleytr,U.B. and Sara,M.		
REFERENCE		Submitted (25-MAR-1998) Zentrum fuer Ultrastrukturforschung,		
AUTHORS		Universitaet fuer Bodenkultur, Gregor Mendelstrasse 33, A-1180		
TITLE		Vienna, Austria		
JOURNML		3 (bases 1 to 3680)		
REFERENCE		Jarosch,M., Egelseer,E.M., Mattanovich,D., Sleytr,U.B. and Sara,M.		
AUTHORS		Direct Submission		
TITLE		Submitted (13-APR-1999) Zentrum fuer Ultrastrukturforschung,		
JOURNML		Universitaet fuer Bodenkultur, Gregor Mendelstrasse 33, A-1180		
REFERENCE		Vienna, Austria		
AUTHORS		Sequence update by submitter		
TITLE		On Apr 13, 1999 this sequence version replaced gi:3025825.		
JOURNML		Location/Qualifiers		
REMARK		1..3680		
COMMENT		/organism="Bacillus stearothermophilus"		
FEATURES		/strain="ATCC12980"		
SOURCE		/db_xref="taxon:1422"		

```

mRNA      /db.xref="ATCC:12980"
           51..3643
           /gene="sbsc"
gene       51..3643
           /gene="sbsc"
misc-feature 84
           /note="alternative transcription start site"
RBS        255..261
           /gene="sbsc"
sig-peptide 273..362
           /gene="sbsc"
           /note="S-layer protein; sbsc"
CDS        /codon_start=1
           /transl_table=11
           /product="surface layer protein precursor"
           /protein_id="AAC12757.1"
           /db_xref="GI:3025826"
           /translation="MDKKAVKLATASAVASAFVAANPHASQATDVATVYSQAKQ
           MKEAYTTSHTTETGQPPDIDVTAAYKAKQATNANAYVYNNKAGAKKDYILDLQ
           ALETVEKANKSGEARVATYIDAINTATKDKMDELKAAVADAKLKADELHKI
           SYELKRTVYILDRVYGOSTRRELRSFEKQAOLRDLTYDITVAMKAREADAVKAG
           NIDKAKALADQVNOYVSKYTDAFKALQCAQDAKAAAYEALTPVESASIDSTSFK
           VFTKPVDKATAIPKNESTLTKETKLPKSEVESSEGLATVTLLYDTLVGKTYFV
           VTSGLKDRAGKEEFTSTNEFTYNNKPVPAITFENPKLPEDSAVDLFKYTVKDAAGNV
           IKSFELEPTESEKLTQCKFIWTTGKSVTVAAYKGTWTTGNTILANEDKAAEVS
           ELKLRDNKEVNTLLANGNAFEDKDNQISSGLTLLAKKDYIGMELTKVAGDTYTE
           ESLNPEVLVAVDGSVTPVPGTALVYVYGVETKIIPVTKANPVLTEIADVDSGV
           VAKGOKATFEKVLKDOYGNKFTGNVMTSDKTEIATVSNANGIGOSEYTVYNGVAE
           GSTTILIKSGTEKREVPVNVAGVPVANOIYKLDGKIDKATESPANDVOLKVA
           VDNANGINVDINDVYITSEATDNTGVYNAKSTANDGTLAKILANLDRKDNQNG
           VLGVTYLTGTVNEVIDITLKATPVYTKKDLIELDANBDLAKILANLDRKDNQNG
           PNVDSAAIPNINEKLOALKSVLSGVSVSPTVIGSVSNVNDLKDASISGLAVKRG
           VTLVLVFNEDSKLAPLAIYVKAAPATODGVTVGLDVGVTGVTGKTKTADKIKSG
           HKLYLAVDSDAIVPAVGTTRNSTKFAENITVGTTEVAANAGQIIIVIVDSDDRVG
           KXTTFVEADLISVADKRTGTATVTPTGNOVYTGKTLIAVSDILANGHKLVAAGSS
           AAAPVKGIAYDTYRTYRTYGTGTEVTSIGTEVDAQSGQHSIIIEVDENCKVYVXQYTI
           TGIQIGTKSAS"
mat_peptide 363..369
           /gene="sbsc"
           /product="surface layer protein"
terminator 3606..3643
           /gene="sbsc"
BASE COUNT 1265 a 694 c 766 g 955 t
ORIGIN
Query Match 13.9%; Score 513; DB 1; Length 3680;
Best Local Similarity 66.6%; Pred. No. 3e-75; Indels 9; Gaps 2;
Matches 767; Conservative 0; Mismatches 375;
QY 1 atgagataggaagaaagctgtgaactagcaacagcaagtctattcagcaagtgatcatt 60
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 273 ATGGACAAAAAGAAAGCAGTCAACAGTCCGCAAGCCGCTGTAGCAGCAGCGCTTTC 332
QY 61 gtgcgtgcaaatcacaacgcttcagaagcggtcagatgtagcaacagtagtaagccaa 120
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 333 GTGCGAGCCAAACCCGATGCTTCCCAAGCAGCAACGACGTGCGAGCGTGTGACGCAA 392
QY 121 gcaaaagcagagttcaaaaagatactactactacagcagcatcagtaagaagaaactggt 180
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 393 GCGAAAAGCGCAAAATGAAGAAGCATACTATACGTACAGCATACGGTAACGGAAACGGGGC 452
QY 181 gaattcccaacattacagatgatatgtcgtgaatacaacaagcgaagaaacatacgcgt 240
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 453 GATTCGCCGACATTAAGATGTATATGCGCGCTTACACAAACAAACAAAGGATATGCA 512
QY 241 gatcgagtagcatagtagtaagaagcaggtgagcggaagaaagacgttacttagctgat 300
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 513 AACGCGGTGCTGCTGTGAACAAAGCCGGCGCGCAAAAAGACGCGTATTTGGCTGAT 572

```

```

QY 301 ttacaaaagaatatgaactacgttttcaaaagcaaaccttaaatctggtcgagctgt 360
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 573 TTGCAAGCATATATACGAAGCATATGATTTTCAAAAGCCAAACCGCAATGGGGCAAGCGCC 632
QY 361 gtagcaacttaacatcgatgcttaacaactatgcaacaacaaatagaagaaatgcgcaag 420
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 633 GTGGCTACGTACTACATCGACCTTACCACTATGGCACAATAATGACAAAATGGCCAAAGA 692
QY 421 ctgaagagctgctgttcaagaacaaagatttagaagaaagcagaacaatactatcaaat 480
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 693 CTCAAAAGCAGCCGTCGATCGATCGAAGATCTGAAAAAGAGAGAGCTTACCAACAAATTT 752
QY 481 ccctatgaataataaactgcacagcagtcatttagatcggtatgtaataaactcgt 540
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 753 TCGTACGAGCTCAAAACGGCGCAAGCATCTTCGACCGGCTCAGCGCAATGACGCC 812
QY 541 gattactctgcctctacatttaagcaaaagcaagcaaaccttcgcagcgttaattat 600
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 813 GAACCTCTCCGTCGACGTTCAAGCCGATGCTCAAGCATTTGCGGACAGATTGATCTAC 872
QY 601 gatattacggttgcaatggaagcgcggaagtaacaagcagctgtgtaacagcaattta 660
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 873 GACATCACCGTTGCCATGGAAGCGCGTGAAGCGCAAGACGCTGTCAAAGCGCAACTTG 932
QY 661 gacaaagctaaagctgtctgtgaatcaatcaatcaatcaatcaatcaatcaatcaat 720
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 933 GCAAAAAGCAGAAAGCCGCGCTTGACCAAGTCACCAATACGTATGCAAGTAACCGATGG 992
QY 721 ttcaaaactgaagctgtctgtgaatcaatcaatcaatcaatcaatcaatcaatcaat 780
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 993 TTCAAAAGCGCAACTGCAAAAAGCAGCAGCAAGACGCCAAGCGCGCGTATGAAGCATTTG 1052
QY 781 accccaagttggaaggttaagtgatgataacaccacaacaaacgttgaataaca 840
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 1053 ACCCGAAGGTTGAGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 1112
QY 841 gcagtcacagtcgaacgaacacataaacttaacacttcagctgtcgaatgaatga 900
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 1113 AACCGCTGACCAACCAACACAGCAATTCCTAAAACCTTCTATTTACTTAAAGATGACT 1172
QY 901 gtaacgtaatactctgtaacatctataaagtggaagcgaataacatcattgccttaac 960
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 1173 GAAA---CTAAACTTATCTCTAAGTCTGTGGAAGTATCGAGAGTGAATTAAGTCTACT 1229
QY 961 aagcgagatgcttcttctatctcaagaagcgaacacatacactacgtgtgatactcaat 1020
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 1230 GTAACTCTTTATGATCATTAAGTGAAGGTTGAAGGTTGAAGGTTGAAGGTTGAAGGTT 1286
QY 1021 ttccgaataataagcagagataaagtagttagtaagaagcattaaagcaaaaatggc 1080
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 1287 ---TTAAAAGATACAGCTGTGTAAGAAATTTGAACATAGCACAACAAATTCACCTTCAAC 1343
QY 1081 gaatttaagaaatgacatcattcaagcttcgaatgaatgctgttagttagtcaagtg 1140
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 1344 AAGCCAGTACACAGCTCTTATACATTCACACTTCAATAAAGTATACCTGAAGATATGCAATT 1403
QY 1141 ttggaactaa 1151
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 1404 GATTGTGACTAA 1414

```

```

RESULT 5
AF228338 3090 bp DNA BCT 24-JUL-2000
LOCUS AF228338
DEFINITION Bacillus stearothermophilus SpsD precursor (spsD) gene, complete cds.
ACCESSION AF228338
VERSION AF228338.2 GI:9188638
KEYWORDS Bacillus stearothermophilus, Bacillus stearothermophilus, Bacteria: Firmicutes, Bacillus/clostridium group; Bacillus/Staphylococcus group; Bacillus.

```


[illegible][illegible]

```

CC FT /note="maltoase-binding protein"
CC FT 1150..1174
CC FT /note="EcoRI adaptor"
CC FT 1174..3195
CC FT misc_feature
CC FT /note="immunogenicity protein"
CC FT 3194..3218
CC FT /note="EcoRI adaptor"
CC FT 3219..3399
CC FT misc_feature
CC FT /note="sequence derived from pMAL-c vector"
XX
FH Key Location/Qualifiers
FH source 1..3399
FH /db_xref="taxon:32644"
FH /organism="unidentified"
FT
FT XX
SQ Sequence 3399 BP; 1577 A; 509 C; 797 G; 516 T; 0 other;

```

```

Query Match 2.1%; Score 77.4; DB 22; Length 3399;
Best Local Similarity 43.0%; Pred. No. 0.0014;
Matches 436; Conservative 0; Mismatches 576; Indels 3; Gaps 1;

QY 84 tgaagcggctacagatgtacacacagtagtaagccaaagcagtcacaaagc 143
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DQ 2064 TGAAGAAAGAAAAGAAAGAAAGTAAACATGAAAGAAATAGAAAAGAGCATGAAGAAAGT 2123

QY 144 atactatcttaacagccctacagtaacggaactgtgaattcccaacattacagatgt 203
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DQ 2124 AATACATACAGACAGAAAAGAAAGAAAGTAAACATGAAAGAAAGAAAGAAAGAAAGTAA 2183

QY 204 atatgtcttaacacaaagcgaataacgtagtcgtatgcgtgtagcattagtataa 263
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DQ 2184 AGAAGTAAATACATGAAAGAAAGAAAGTAAACATGAAAGAAAGAAAGAAAGTAAAGT 2243

QY 264 agcaggtgctgcgaataacagcgtctacttagctgattacaaagaatataagaaactta 323
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DQ 2244 AACACATGAAAGAAAGAAAGAAAGTAAATACATGAAAGAAAGAAAGAAAGTAAATACA 2303

QY 324 cgtttcaagcaaaccttaactctgcgaagctcgtgtacaaacttaccatgactta 383
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DQ 2304 TGAAGAAAGAAAGAAAGAAAGTAAACATGAAAGAAAGAAAGAAAGTAAACATGAAAG 2363

QY 384 caactatgcacaaatagacgaatctgcgcgaagtagagctgtgttcaagcaaa 443
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DQ 2364 AGAAGAAAGAAAGTAAACATGAAAGAAAGAAAGAAAGTAAACATGAAAGAAAGAAAG 2423

QY 444 agatttagaataacagcaatactatcacaataatccctatgaaatataaactgcac 503
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DQ 2424 AGAAGAAAGTAAACATGAAAGAAAGAAAGTAAACATGAAAGAAAGAAAGTAAAGT 2483

QY 504 agctatttagatcggtatctgttaaaacactcgtgatttactctgctctacattaa 563
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DQ 2484 AACACATGAAAGAAAGAAAGAAAGTAAACATGAAAGAAAGAAAGAAAGTAAACATG 2543

QY 564 agcaaaagcacaagaactctgcgcgaagcttaattatgatttaccgttgcataagc 623
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DQ 2544 AGAAGAAAGAAAGTAAACATGAAAGAAAGAAAGTAAACATGAAAGAAAGAAAGTAA 2603

QY 624 gcgcgaagttaacagcgtctgtgaagcgcgcaatttagaacaagctaaagctgtgtta 683
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DQ 2604 AGAAGAAAGAAAGTAAACATGAAAGAAAGAAAGTAAACATGAAAGAAAGAAAGTAA 2663

QY 684 tcaactatcaatacttaccaaagtagaagatcttccaaactgaactaaacgaagct 743
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DQ 2664 ACTAACACATGAAAGAAAGAAAGTAAATCATG--AAGAAAGAAAGAAAGAAAGTAA 2720

QY 744 agcgaataaacattagatgcagatgaagctgcgttactccaaagctgaagctgaag 803
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DQ 2721 TGAGGAAGAAAGAAAGAAAGAAAGTAAAGAAAGAAAGAAAGTAAAGTAAAGTAAAG 2780

QY 804 tgcgatttaacactcaaaagctgttgaatttaacagcagtagtgaagcgaagcact 863
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

DQ 2781 AGACGAAGAAAGAAAGAAAGTAAAGAAAGTAAAGAAAGAAAGTAAAGAAAGTAAAG 2840
QY 864 aaattcaacttcacgtcgtcgaatgaagatacagtaacgttaactgaactat 923
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DQ 2841 AGAAGAAAGAAAGTAAAGAAAGAAAGAAAGTAAAGAAAGTAAAGTAAAGTAAAGTAA 2900
QY 924 ctataagctgacgttaacattccttgccttaatacgcgagatgtttcttatctac 983
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DQ 2901 AACACATGAAAGAAAGAAAGTAAACATGAAAGAAAGAAAGTAAACATGAAAGTAA 2960
QY 984 agacggaaatactatcactgtggtatgtcttaccctcattcgaaataatacaggtataa 1043
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DQ 2961 AGAAGAAAGAAAGTAAACATGAAAGAAAGAAAGTAAACATGAAAGTAAAGTAAAG 3020
QY 1044 agtagtgaattgaattaaagcaaaatgccaagaatttaagaaagatgca 1098
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DQ 3021 AAAAGTAAACATGAAAGAAAGAAAGTAAACATGAAAGAAAGTAAAGTAAAGTAA 3075

```

```

RESULT 8
AP002460 111554 bp DNA PIN 19-JUL-2000
LOCUS AP002460
DEFINITION Arabidopsis thaliana genomic DNA, chromosome 3, BAC clone:PID9.
ACCESSION AP002460
VERSION AP002460.1 GI:8347620
KEYWORDS
SOURCE
ORGANISM
Arabidopsis thaliana (strain:Columbia) DNA, clone_lib:IGF BAC
clone:PID9.

```

```

REFERENCE
AUTHORS Nakamura,Y.
TITLE Structural Analysis of Arabidopsis thaliana Chromosome 3. III
JOURNAL Unpublished (2000)
REFERENCE 2 (bases 1 to 111554)
AUTHORS Nakamura,Y.
TITLE Direct Submission
JOURNAL Submitted (06-JUN-2000) to the DDBJ/EMBL/Genbank databases.

```

```

Yasukazu Nakamura, Kazusa DNA Research Institute, Department of
Plant Gene Research, 1532-3, Yana, Kisarazu, Chiba 292-0812, Japan
(E-mail: ynakamu@kazusa.or.jp, Tel: 81-438-52-3935,
Fax: 81-438-52-3934)
Address for correspondence: kase@kazusa.or.jp
For the latest information on annotation of this clone, please see
http://www.kazusa.or.jp/kaos/cgi-bin/sgd_graph.cgi?c=PID9
Genes with similarity to proteins in the databases are described in
'product' or 'note' qualifiers. Genes that have no significant
protein similarity are described as 'unknown protein'.
The software programs used to predict genes include: Grail
(Informatics Group, Oak Ridge National Laboratory,
http://compbio.ornl.gov/Grail-1.3/),
GENSCAN (Chris Burge, MIT, http://CCR-081.mit.edu/GENSCAN.html),
NetGene2 (S.M. Hebsgaard, et al., CBS, Technical University of
Denmark, http://www.cbs.dtu.dk/services/NetGene2/) and
SplicePredictor (Volker Brendel, Stanford University,
http://gremlin.zool.jastate.edu/cgi-bin/sp.cgi).

```

```

Genes encoding tRNAs are predicted by tRNAscan-SE
(Sean Eddy, Washington University School of Medicine, St. Louis,
http://genome.wustl.edu/eddy/tRNAscan-SE/).
This sequence may not be the entire insert of this clone. It may be
shorter because we remove overlaps between neighboring submissions.
The 5' clone is T889 and the 3' clone is T789.
Location/Qualifiers
1..111554
/organism="Arabidopsis thaliana"
/strain="Columbia"
/db_xref="taxon:3702"
/chromosome="3"
/clone="PID9"
/clone_lib="IGF BAC"
complement(join(4747..5195,5297..5754,5827..6284,

```

```

FEATURES
Source

```

CDS

```

6457. .6803,6864. .7079,7216. .7286,7326. .7810,7914. .8253,
8311. .8737,9083. .9347))
/note="gene_id:F1D9.1"
/codon_start=1
/pseudo
/evidence-not_experimental
/product="helicase-like protein"
join(11589. .11736,11836. .12019,12305. .12497,12774. .12896,
13021. .13080)
/note="gene_id:F1D9.2"
/codon_start=1
/evidence-not_experimental
/product="ATP-dependent CLPB protein-like"
/protein_id="BAA37091.1"
/db_xref="GI:8777573"
/translation="MLEQILHRYIADIDVDSYADAIRCSKAGISDPNRIASFMTN
GPISVDSQVGSVGYGVGDEGSKSTEVRKRPYSVVOEDIEKPHHPVFSILQLLD
DGRITDSHGSALHILETIRNNEDEKNAVEMKQOVELARKTERKPMRNIIDYIS
QPLNSVYSKIVELOVQVQMRKRLLEQPNRLITLNLKGFPAENDTILIDADQPF
NKLVIKKMENAHRKON"
complement(join(19018. .19216,19317. .19390,19600. .19746))
/note="contains similarity to unknown protein"
gb|AAC97228.1
gene_id:F1D9.3"
/codon_start=1
/evidence-not_experimental
/protein_id="BAA97092.1"
/db_xref="GI:8777574"
/translation="MAKKEKRVVREEDVYKRMKLTIMPLYNESLHQQLALFSGYP
STLIKINYMKEEMKMHKQEFIASRSGSFSTGATIPMCARINIKDPKRGIAVL
PGKILITLTVSKSRKSGIKDKLIIIVIVKLVMP"
join(21817. .22004,22075. .22143)
/note="contains similarity to ATP-dependent CLPB protein"
gene_id:F1D9.4"
/codon_start=1
/pseudo
/evidence-not_experimental
/complement(22489. .23812)
/note="contains similarity to retroelement pol
polyprotein"
gene_id:F1D9.5"
/codon_start=1
/pseudo
/evidence-not_experimental
join(24404. .24543,24659. .24879)
/note="contains similarity to ATP-dependent CLPB protein"
gene_id:F1D9.6"
/codon_start=1
/pseudo
/evidence-not_experimental
complement(25073. .25379)
/note="gene_id:F1D9.7"
/codon_start=1
/pseudo
/evidence-not_experimental
/product="retroelement pol
polyprotein-like"
join(30097. .30190,30265. .30551,30614. .31510,31758. .32036)
/note="gene_id:F1D9.8"
/codon_start=1
/pseudo
/evidence-not_experimental
/product="ATP-dependent CLPB protein-like"
join(40639. .40982,41779. .42326,42393. .42794,42903. .42938,
43020. .43142,43226. .43333,43425. .43481,43588. .43650,
43733. .43973,44026. .44107,44503. .44674,44813. .44940,
45043. .45198)
/note="gb|AAD25852.1
gene_id:F1D9.9
strong similarity to unknown protein"
/codon_start=1
/pseudo
/evidence-not_experimental
join(46889. .47780,47984. .48332,48695. .48783)

```

```

/note="gene_id:F1D9.10
pir||T01475
similar to unknown protein"
/codon_start=1
/pseudo
/evidence-not_experimental
join(51111. .51234,51470. .52398,52474. .52672,52762. .53360,
53433. .53558,53717. .53809,53908. .54010,54110. .54210,
54472. .54531,54627. .54728,54795. .55047,55142. .55226,
55334. .55637,57320. .57389,57601. .58187)
/note="gene_id:F1D9.11
pir||T02399
strong similarity to unknown protein"
/codon_start=1
/pseudo
/evidence-not_experimental
complement(join(58782. .58995,59066. .59150,59231. .59498,
60276. .61922,61970. .62259,62305. .62482))
/note="gene_id:F1D9.12"
/codon_start=1
/pseudo
/evidence-not_experimental
/product="maize transposon MudR protein-like"
complement(join(63866. .64008,64160. .64226,64346. .64414,
64504. .64620,64743. .64790,64956. .65053,65183. .65264,
65486. .65566,65677. .65718,65854. .65952,66106. .66144))
/note="contains similarity to human diaphanous 1
'gene_id:F1D9.13"
/codon_start=1
/evidence-not_experimental
/protein_id="BAB01984.1"
/db_xref="GI:9294133"
/translation="MLSKVITPLPDLMSVLADSEYIDVDQVNLKRCPTKEEARL
LKGTGKKEIDGRCDFLELDKVPKVEIKURFSEFIQHSQVRSSTKLRKQITL
SLGNALNHGTARSAIGFHLDSILKLTDRSRRNKRLMHGLCKGVAELKPIGLNFP
KDMIQLYLAEEMQATSGLEKVVQEFSTASTDCQISKHFPMNLKEFLSYAEQVRSI
ASLSTVGSADALATYGEDPARVPEQVYSTLQNFVRLFVRSHENCKQVEKKR
'AKRAENKLRKKYNNEN"
complement(66524. .66974)
/note="gb|AAD23008.1
gene_id:F1D9.14
similar to unknown protein"
/codon_start=1
/pseudo
/evidence-not_experimental
complement(join(68701. .69561,69697. .70591,70640. .70667,
71526. .72709,72898. .73300,73370. .73533,73687. .74594))
/note="gene_id:F1D9.15"
/codon_start=1
/pseudo
/evidence-not_experimental
/product="retroelement pol
polyprotein-like"
complement(75178. .76058)
/note="gb|AAD23008.1
gene_id:F1D9.16
similar to unknown protein"
/codon_start=1
/pseudo
/evidence-not_experimental
76515. .80547
/note="emb|CA875469.1
gene_id:F1D9.17"
/codon_start=1
/pseudo
/evidence-not_experimental
/product="copia-type reverse transcriptase-like protein"
complement(join(81294. .81521,81620. .82285))
/note="gb|AAD23008.1
gene_id:F1D9.18
strong similarity to unknown protein"
/codon_start=1
/pseudo
/evidence-not_experimental

```

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS


```
QY 341 cttaactctgagagctcgtctagcaacttaccatctctcaactatgcacacaat 400
    || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3006 AAGAGTAAAGAAAAGATACCGAAGCAAGATTAAGAGATAGAACAGAAAAGAAA 3065
QY 401 tagacgaatgcgcgaagctagagctgctgtctcaagcaaaagattagaagaagcg 460
    || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3066 AAGAAAGAAATAAAGAGTTAAAGAAAAAGATCCGAAAAACAAGATTAAGTATGAGC 3125
QY 461 aacaaat---acatacacaattcctatgaatataaactgcgcagatcatttagac 517
    || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3126 AAGAAATTAATTAATAGAAATTAATAAAGAACTTAAAAAGATTAATAAAGAAATA 3185
QY 518 gctatatgtataaacaactcgtatctactcgcctcatcttaagcaaaagcccaag 577
    || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3186 ATTAATAATTAATAAACAAGATTAATGTATGTACAAAGAAATTAATGAACGAGATTTCGAA 3245
QY 578 aacttcgcgacgcttaattatgatattaccgttcgaatgaagcgcggaagtacaag 637
    || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3246 AAGATTAACCCCAACAAGATTAAGGTATGAAACAAAGAAAAAGAAAAAGAAAGTAAAG 3305
QY 638 agcgctgtgaagcaggaacttagacaaagcctaagctgctgtgttcaatcaatcaat 697
    || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3306 AAAAAAGAGAGTAAAGAAAAAGAAAGTAAAGAAAAAGAAAGTAAAGAAAAAG 3365
QY 698 acttaccaaagtaacagatgtcttcaaaactgaactaacagagtagcgaaagaagcat 757
    || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3366 AAGAGTAAAGAAAAAGATTAACCGAAGCAAGATTAAGAGATAGAACAAAGAAAAAGAAA 3425
QY 758 tagatgcagatgaagctgcgcttactccaagaagtgaagtgtaagtgcatgaactac 817
    || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3426 AAGAGAAAGTAAAGAAAGTTAAAGAAAAAGATACCGAAAAACAAGATTAAGTATGAGC 3485
QY 818 aaaaaaagctgttaattaaacagcgtaccgtagaagcgaacactaaattacaactt 877
    || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3486 AAGAAATTAATTAATAGAAATTAATAAAGAAAGTAAAGAAAGTAAAGAAAAAGAAATA 3545
QY 878 cagctgtctcaaatgaagatacagataaactgtacgtatctataagtagagcg 937
    || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3546 ATTAATAATGAATAAACAAGATTAATGTATGTACAAAGAAATTAATGAACGAGATTTAAAG 3605
QY 938 gtaacattccattgcccttaatacagcagatgtcttcttattctacagcggaataacta 997
    || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3606 AAAAAAGATACCCCAACAAGATTAAGGTATGAAACAAAGAAAAAGAAAAAGAAAGTAA 3665
QY 998 tcaacgtgtagtcttcaactcattccgaaataatacgaagataaagtagtagttaag 1057
    || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3666 AAGAAAAAGAAAGATTAAGAAAAAGAAAGATTAAGAAAAAGAAAGAAAGTAAAGAAA 3725
QY 1058 gttataagacgaataatgcgaagaatttaagaaga 1094
    || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3726 AAGATTAACCGAAAGCAAGATTAAGATTAAGATTAAGATTAAGATTAAGATTAAG 3762
```

RESULT 10
AC024253
LOCUS AC024253 179553 bp DNA HTG 07-JUL-2000
DEFINITION Homo sapiens chromosome 6 clone RP11-758C19, WORKING DRAFT
ACCESSION AC024253.3 GI:8568958
VERSION AC024253.3 GI:8568958
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 179553)
AUTHORS Waterston,R.H.
TITLE The sequence of Homo sapiens clone
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 179553)
AUTHORS Waterston,R.H.
TITLE Direct Submission

JOURNAL COMMENT

Submitted (28-FEB-2000) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA
On Jun 16, 2000 this sequence version replaced gi:725339.

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: <http://genome.wustl.edu/gsc/index.shtml>
Project Information -----
Center project name: H_NH0758C19
----- Summary Statistics -----
Sequencing vector: M13, 100%
Sequencing vector: Plasmid, 0%
Chemistry: Dye-terminator Big Dye, 0% of reads
Chemistry: Dye-terminator Big Dye, 0% of reads
Assembly program: Phrap, version 0.990319
Consensus quality: 175235 bases at least Q40
Consensus quality: 176522 bases at least Q30
Consensus quality: 177171 bases at least Q20
Insert size: 198000; agarose-fp
Insert size: 178453; sum-of-contigs
Quality coverage: 4477.53 in Q20 bases; agarose-fp
Quality coverage: 4.99 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently consists of 12 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.
*
* 1
* 2275 2274: contig of 2274 bp in length
* 2375 2374: gap of unknown length
* 4602 4601: contig of 2227 bp in length
* 4702 4701: gap of unknown length
* 8521 8520: contig of 3819 bp in length
* 8621 8620: gap of unknown length
* 16435 16434: contig of 7814 bp in length
* 16535 16534: gap of unknown length
* 24089 24088: contig of 7554 bp in length
* 24189 24188: gap of unknown length
* 32983 32982: contig of 8695 bp in length
* 32984 32983: gap of unknown length
* 43009 43008: contig of 10026 bp in length
* 43010 43009: gap of unknown length
* 43110 43109: gap of unknown length
* 52415 52414: contig of 9305 bp in length
* 52415 52414: gap of unknown length
* 78545 78544: contig of 26030 bp in length
* 78545 78544: gap of unknown length
* 107100 107099: contig of 28455 bp in length
* 107100 107099: gap of unknown length
* 133817 133816: contig of 26518 bp in length
* 133818 133817: gap of unknown length
* 179553 179552: contig of 45736 bp in length.
*
FEATURES
source
1..179553
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="6"
/clone="RP11-758C19"
1..2274
/note="assembly_name:Contig6"
2375..4601
/note="assembly_name:Contig7"
4702..8520
/note="assembly_name:Contig8
clone_end:77
vector_side:right"
misc_feature 8621..16434
/note="assembly_name:Contig9"
16535..24088
misc_feature


```
/product-"dUTPase"
/protein_id-"AAA46131.1"
/db_xref-"GI:330983"
/translation-"MPYKPEIYREPOTFYITSPARASNTQLNHNILYKAGCV
ITVSGIIFKVFSAFILDYKSAKSIKCHTGLIDPEOGELKLVINKREDITDITN
DLRVATVAEVVGVKFLHDYSDLPKPSYDAGEFDLYLPDVAVKPVPKRYSPNCCP
AOLKSKPVLFGRSGLAKGLTIKVRMONGQLIIFYNTRKSOITATYATRIAGVFM
HKHLPTLRLKPTMHLSENIKYSARVSPDIDKTFVQDEKLVSSSDTSDSQMR
GDAGLGSSGIM"
5761..5766
/feature="EDRF3 ORF"
/note="EDRF3 ORF"
complement(5793..5798)
/note="EDLF1 ORF"
complement(5797..6399)
/note="first methionine codon in the EDLF1 ORF"
/codon_start=1
/protein_id-"AAA46132.1"
/db_xref-"GI:330984"
/translation-"MLQWLHSCMPFGKRYRQKYPVLDITSSCPDRMKRIEIAOPGV
FVGDILQNSDSDASLRQAYLLAVOLNITDYLKRFDEASVPSKSVQVOTLRKSV
RNIWNTMLSMAGVGTIDDAIKTLIDPKRAGESTALIMELKLAIVYVDSKNAKE
INNTLSAHEKQIIVNSEVPLIECETTLAERTTPAVSI"
6378..8885
/feature="EDRF4 ORF"
/note="first methionine codon in the EDRF4 ORF"
/codon_start=1
/function="required for DNA replication"
/protein_id-"AAA46133.1"
/db_xref-"GI:330985"
/translation-"MNEATEAMPKFRKYLEFADGSAEITDILGTDITNATYSLHN
CIYPTVEKIVILCLPAKPGGDKCLEFOLHIDTELAIFLFTYKPLKANDLHKY
IDFKARKNFILDIISTNPKSPKTHNDISKIYWEPAKVNLSRKIKYSSPYM
MTTFGSEFVPLTALIFYFEFOHNTIYIFHLSFSGEKGKGTSLATTEBELGV
CSTSDYIKTAPAFIVYCHIKIARLSQADISDTLBDQMLSNODLVHYIYSF
OCLANDITIKYSHLNSDNIHFVPEYEVASQSIDENFRKDMFTYKSTLYLTITIK
CIHLPDLIGYAPDCTSEFYVAGQSNVHNLNVINTPHINISEDNLGLDLAID
STFVNDNLKDSYENSOQVRYRCEFLNKFYIYONDILKNVSTDYMLPQEWY
LKDEITSNISYKETFSTMLTLDQLKISREYFNPRLKVPVNLVDLHITSEHI
DEIYMLCTIRSLIETTLQGLPDIDPHYVEFEKSTCKPENMIDNKEIKCYCTK
IGFRITPLPAGVILGSPVSYFNILNRTIKIDKILAMPILIMEDIDPPDVATY
KRCVRIPIHTIVNNSGRLERILKLVCHPHVDNKLQYVDSFDINLLYSHANEK
KOLKAVYDIADINENEILOKAQAOIPQTHNNAVERIESHSNISTDWAEEFMRLE
LIKLYSEKVSQFYHVSFAASTGNIKIISLGSFSCINFHRILKLTOSVRIEFLML
TPDNCVTTLTLMQCFASCNKICIAHMSVVPIDK"
8881..8886
/feature="EDRF4 ORF"
/note="EDRF4 ORF"
complement(9074..9079)
/note="EELF5 ORF"
<9082..910320
/note="5' end of coding region for EYF1 undetermined."
/codon_start=1
/product-"transactivator"
/protein_id-"AAA46123.1"
/db_xref-"GI:330986"
/translation-"YLGISDDPDSDDSSSDSESDSTSPQIMKSDVYMASPSTPEP
SDVASTSNLKREROSPITWEHOSPLSRVYRSPSPMFGPRISLSTSRKST
NADRYREAAORRSPRPKPYSHRNGGLRPPAPLLKFLPISILPKSGEKIF
LVPSPLOEAKETNDKYLAAMORAMHVPYSSQUTANILPYKFKPLISYRTPN
YIYVYSRKETITASNLCTVAFLDESICMGQOITLNDITSENGDITLDSYLLS
QIVHKIKMLPFCHELMQTPDQIVKQVCYLLASNNRLIDAVRYLQTSYIKSPYLL
AAVCLPAIICITKNETOLYSHCMRLIKREYRGDVNMLIHESILTOHLNKPSTCATY
TRAIVGTAKNTGFLFPLQ"
10425..10430
/feature="EYF1 ORF"
/note="EYF1 ORF"
complement(10679..11752)
/note="first methionine codon in the EELF5 ORF"
/codon_start=1
/protein_id-"AAA46134.1"
/db_xref-"GI:330987"
/translation-"MIOEMQWVNLIRASCIAGAAPFICPFKSIPTFSITSL
QISITFVNLGAVHTILCITYITFKENSYITFALGIACFILMMLFTLSHMAAPLY
TLIFVNLKICLSLWPIITDYIVLCPEITTYKIELGLNMLLYVWARMAPLY
IFMCPFVLELGMGVALKNFHEHPVENLITCKPIETAKNKYKTKGTEVNNKLIVHN
IGAVFLLAVERASVSVVORLDIFVGMQNTYLLFLVSMLCOMFSLPSNAICVLET
```

```
FAVYIIVHLIDKLPISIGGLVIVILLMQCAIGQIEIIRTKLAGVGGPKLCL
FLCTVCNIVSVALLCCNKSQSL"
complement(11147..11152)
/note="EELF4 ORF"
complement(11749..12855)
/note="first methionine codon in the EELF4 ORF"
/codon_start=1
/function="required for DNA replication"
/product-"DNA-binding protein"
/protein_id-"AAA46135.1"
/db_xref-"GI:330988"
/translation-"MADTQTFHSYGVGFKPCVLOEHTKIYNHAKRKGDIQIEGL
TSSPLRIASSIGEGILSEKINNVSEVENCAMPENVSPTRNISPGNPLTYREL
FGCNVKTATIMSNRSNGKPEPRKIKAKLCCDKVSTRTTRITPAIILFVENVSTV
LSRVYLSKRTAMLOKMLREHAKHTVYVMNNTAVILLYEDSTKYVDHIYDMS
LCNPSKDMGVQNDVORVNTALISALGLCKIGVVPCKIKFESEVLEHCGTPVK
LGSVDTSLQVTLFCTPNLADTQAVADYDPIISASDLITADPADTOEPLTSSNCD
SEVENSPLPLKRRKRENDINKSKRIKIAFNPLI"
complement(11774..11779)
/note="EELF2 ORF"
complement(11774..11779)
/note="EELF2 ORF"
complement(11774..11779)
/note="EELF3 ORF"
complement(12966..13883)
/standard_name="ribonucleotide reductase beta subunit"
/EC_number="1.17.4.1"

Query Match 1.7%; Score 64; DB 73; Length 43658;
Best Local Similarity 46.3%; Pred. No. 0.14;
Matches 211; Conservative 0; Mismatches 245; Indels 0; Gaps 0;

polyA_signal
polyA_signal complement(11774..11779)
CDS complement(12966..13883)
/standard_name="ribonucleotide reductase beta subunit"
/EC_number="1.17.4.1"

polyA_signal
polyA_signal complement(11147..11152)
CDS complement(11749..12855)
/note="first methionine codon in the EELF4 ORF"
/codon_start=1
/function="required for DNA replication"
/product-"DNA-binding protein"
/protein_id-"AAA46135.1"
/db_xref-"GI:330988"
/translation-"MADTQTFHSYGVGFKPCVLOEHTKIYNHAKRKGDIQIEGL
TSSPLRIASSIGEGILSEKINNVSEVENCAMPENVSPTRNISPGNPLTYREL
FGCNVKTATIMSNRSNGKPEPRKIKAKLCCDKVSTRTTRITPAIILFVENVSTV
LSRVYLSKRTAMLOKMLREHAKHTVYVMNNTAVILLYEDSTKYVDHIYDMS
LCNPSKDMGVQNDVORVNTALISALGLCKIGVVPCKIKFESEVLEHCGTPVK
LGSVDTSLQVTLFCTPNLADTQAVADYDPIISASDLITADPADTOEPLTSSNCD
SEVENSPLPLKRRKRENDINKSKRIKIAFNPLI"
complement(11774..11779)
/note="EELF2 ORF"
complement(11774..11779)
/note="EELF2 ORF"
complement(11774..11779)
/note="EELF3 ORF"
complement(12966..13883)
/standard_name="ribonucleotide reductase beta subunit"
/EC_number="1.17.4.1"

Query Match 1.7%; Score 64; DB 73; Length 43658;
Best Local Similarity 46.3%; Pred. No. 0.14;
Matches 211; Conservative 0; Mismatches 245; Indels 0; Gaps 0;

polyA_signal
polyA_signal complement(11147..11152)
CDS complement(11749..12855)
/note="first methionine codon in the EELF4 ORF"
/codon_start=1
/function="required for DNA replication"
/product-"DNA-binding protein"
/protein_id-"AAA46135.1"
/db_xref-"GI:330988"
/translation-"MADTQTFHSYGVGFKPCVLOEHTKIYNHAKRKGDIQIEGL
TSSPLRIASSIGEGILSEKINNVSEVENCAMPENVSPTRNISPGNPLTYREL
FGCNVKTATIMSNRSNGKPEPRKIKAKLCCDKVSTRTTRITPAIILFVENVSTV
LSRVYLSKRTAMLOKMLREHAKHTVYVMNNTAVILLYEDSTKYVDHIYDMS
LCNPSKDMGVQNDVORVNTALISALGLCKIGVVPCKIKFESEVLEHCGTPVK
LGSVDTSLQVTLFCTPNLADTQAVADYDPIISASDLITADPADTOEPLTSSNCD
SEVENSPLPLKRRKRENDINKSKRIKIAFNPLI"
complement(11774..11779)
/note="EELF2 ORF"
complement(11774..11779)
/note="EELF2 ORF"
complement(11774..11779)
/note="EELF3 ORF"
complement(12966..13883)
/standard_name="ribonucleotide reductase beta subunit"
/EC_number="1.17.4.1"

RESULT 14
HSGEND 112930 bp DNA VRL 29-SEP-1992
LOCUS Herpesvirus saimiri complete genome DNA.
DEFINITION X64346
ACCESSION X64346.1 GI:60320
VERSION complete coding region.
KEYWORDS Saimiriine herpesvirus 2.
SOURCE Saimiriine herpesvirus 2
ORGANISM Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
```


THIS PAGE BLANK (USPTO)